

Chapter Number

Cellular-Automata-Based Simulation of the Settlement Development in Vienna

Reinhard Koenig¹ and Daniela Mueller²

¹*Bauhaus-University Weimar, Faculty of Architecture, Chair Computer Science in Architecture,*

²*TU Vienna, Department of Spatial Development, Infrastructure & Environmental Planning, Centre of Regional Science*

¹*Germany*

²*Austria*

1. Introduction

The structure and development of cities can be seen and evaluated from different points of view. By replicating the growth or shrinkage of a city using historical maps depicting different time states, we can obtain momentary snapshots of the dynamic mechanisms of the city. An examination of how these snapshots change over the course of time and a comparison of the different static time states reveals the various interdependencies of population density, technical infrastructure and the availability of public transport facilities. Urban infrastructure and facilities are not distributed evenly across the city – rather they are subject to different patterns and speeds of spread over the course of time and follow different spatial and temporal regularities.

The reasons and underlying processes that cause the transition from one state to another result from the same recurring but varying pronounced hidden forces and their complex interactions. Such forces encompass a variety of economic, social, cultural and ecological conditions whose respective weighting defines the development of a city in general. Urban development is, however, not solely a product of the different spatial distribution of economic, legal or social indicators but also of the distribution of infrastructure. But to what extent is the development of a city affected by the changing provision of infrastructure?

As Lichtenberger (1986, p. 154) already notes, urban structures have often been characterized by the development of technical and socio-cultural infrastructure systems. New buildings erected away from existing roads, should meet certain conditions in terms of their accessibility and waste disposal ("Denkschrift über Grundsätze des Städtebaues," 1906, p.5 ff.). Similarly, one can observe that in the past the development of urban quarters followed the characteristic expansion measures arising resulting from the requirements and extension of road, transport and technical infrastructure systems such as the sewage system ("Denkschrift über Grundsätze des Städtebaues," 1906).

In many European cities, including Vienna, vast infrastructural expansion took place during the Wilhelminian period – the so called "Gründerzeit" – particularly with regard to underground town planning. The network of technical infrastructure, especially sewage,

lighting, gas and water networks was extensive. The development of public transport systems, most notably the tram, advanced rapidly during this period too. Urban structures were influenced considerably by the routing of urban supply and waste disposal networks. During the "Gründerzeit", urban design principles adhered to a hierarchical progression from the centre to the periphery of the city. The centre was most well-equipped and enjoyed the greatest benefit of technical infrastructure. With increasing distance from the centre, the provision of supply and waste disposal systems in the outlying quarters became less extensive. Later, as new suburban districts began to be built, these were equipped with the respective technical infrastructure as part of the building measures. Only later were the older suburbs finally connected to the newer technical infrastructure.

The attraction of an urban quarter as a residential district depended on the degree and variety of technical infrastructure available (Behrens, 1971; Weber, 1909). The supply of urban areas with water, sewage systems and energy influences the provisions and attractiveness of a quarter considerably and in some cases may even create the necessary conditions for residential use in the first place. In addition, an urban quarter's connection to public transport networks determines its accessibility and with it the possibilities of interchanges between different locations. The choice of location can therefore be considered as a result of the analysis and evaluation of various criteria (such as situation and availability of facilities).

Applied to a simulation model the local conditions that characterise the urban development correspond to endogenous and exogenous control parameters. By validating the following simulation model using the historical development of the city of Vienna as a basis, it is possible to derive initial conclusions concerning the driving forces and the abstract configuration of a society from the settings of the parameters.

2. From urbanism to suburbanisation

In the development of a city, one can observe phases of growth and shrinkage. A pattern emerges with growth phases following shrinking phases and vice versa. The simplified model of cyclic phases of urban development assumes that the city has a life cycle (Dangschat, 2007). With the help of such models, it is possible to show the centrifugal processes of suburbanism and desurbanism that set in after the initial centralising processes of urbanism. The different urban development cycles and their interaction are shown schematically in the model of cyclic phases of urban development by Van den Berg et al. (1982) (see Fig. 1).

2.1 Urbanism

The European city of the 19th century is a centralised system where the city centre has the best accessibility, the highest prestige value and the most expensive land prices. The social and monetary capital is concentrated in the city centre – furthermore the city centre represented the overall social, political and economic balance of power. The city is experiencing a phase of urbanism (see Fig. 1) – advantages resulting from the agglomeration of the city are used in many different ways, e.g. to increase the concentration of workers and demand, which in turn stimulates an intensive flow of migration from the rural areas to the cities. However, the cities are not prepared for this huge influx of population – as a consequence, public transport systems were lacking, the working hours long, the incomes low, the quickly built housing units small and poorly equipped. The cities had to extend

public transport systems, supply and waste disposal facilities and social infrastructure within a short time. With the development and construction of rail-based public transport systems, the axis-bound growth along these new routes expanded rapidly, followed by the widening of the urban development radius. However, the city centre remained the location with the best accessibility (Fassmann, 2005, p. 33 and p.104; Maier & Tödting, 2002, pp. 162 ff).

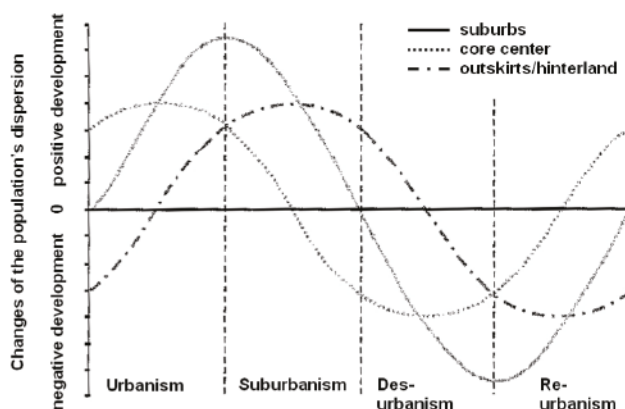


Fig. 1. Model of the cyclic phases of urban development [from Fassmann (2005, p. 105), source: Van den Berg et al. (1982)]

2.2 Suburbanisation

The settlement of the areas between the axes only began after private transport, cars, became widely affordable and available. This led to a rapid increase in the area occupied by the city and a decrease in the accessibility of the inner city due to a lack of parking spaces and restricted traffic areas. The expansion of the public transport systems, the rising incomes and the high densities in the city centre also contributed to the growth of the city, especially on the outskirts. Thus the population density in the city centre began to decline. In contrast, functions that thrive on agglomeration such as special trade, offices and services settled in the city centre, increasingly displacing residential and industrial uses. A cyclic phase of suburbanisation therefore follows the cyclic phase of urbanism (see Fig. 1).

Suburbanisation can be understood as the relocation of land use and population out of the heart of the city, rural areas or other metropolitan areas into the urban hinterland (Friedrichs, 1981). With the expansion of road networks and the widespread availability of cars as a means of mass transportation for the population it became possible to settle on the outskirts of the city and surrounding areas and still have access to the advantages of the city. The city centre can be reached within a reasonable amount of time, so that settlement in areas with more living space and lower prices became increasingly attractive. The urban realm continues to grow, particularly in the suburbs and at the expense of the city centre. Companies that are not dependent on the agglomeration advantages of the city centre relocate their offices to the outskirts, where more space is available at lower prices. Moreover, as a consequence of suburbanisation sufficient manpower and demand is available in the suburbs. Increasing mobility, newly-built privately-owned homes and the expansion of route networks such as public transport, communication and infrastructure lead to centrifugal, decentralised development (Läpple, 2003)

2.3 De-suburbanisation and Re-urbanism

As the city spreads, the travel distances increase and with it the traffic load and the demand for greater traffic capacity. Due to the relatively low population density in the suburbs, infrastructure-intensive public transport systems are uneconomical, a further reason why an increasing proportion of traffic is dominated by private means of transport. As private means of transport expand, the accompanying pollution has a negative effect on the quality of living in many places, so that more and more inhabitants of the city move to the outskirts. The costs of public transport facilities burden the budget of the city, but because more and more of the affluent population has migrated to the outlying regions, the towns find themselves increasingly in a financial bottleneck (Fassmann, 2005, p. 104; Maier & Tödtling, 2002, p. 163). After the cyclic phase of suburbanisation, a new cyclic trend of urbanism sets in, so-called de-suburbanisation and re-urbanism (see Fig. 1). During de-suburbanisation, cities in the surroundings experience a cyclic phase of urbanism while the respective outlying regions stagnate, followed by a cyclic phase of re-urbanism as activity in the city centres increases.

3. Centripetal and centrifugal forces

The hidden forces described above can be loosely summarised into two opposing forces – a centripetal and a centrifugal force (Krugmann, 1996). This chapter describes how they have been derived. Centripetal forces describe centralising forces, which express the advantages and needs of a dense population. Centrifugal forces describe the advantages of decentralised locations, which are primarily the generous availability of space and the absence of polluting emissions.

Myrdal's (1957) centre-periphery model considers, on the one hand, the emergence of centres and peripheries as a result of deprivation or suction effects, so-called "backwash effects" or centripetal forces. These include effects that arise from the attraction of a dynamically-developing city centre, such as the migration of population or production factors from the rural areas. These effects occur primarily during the cyclic phase of urbanism or re-urbanism (see sections 2.1 and 2.3). On the other hand, the centre/periphery structure is a product of so-called "spread effects". They are also described by the centrifugal forces and express the effects of an expanding centre, such as can be observed in the cyclic phases of suburbanisation or de-suburbanisation (see sections 2.2 and 2.3).

Centripetal and centrifugal forces lead to concentrated or dispersed settlement patterns. The key determining factors are the interdependencies between the location decisions of companies, households and the public authorities. The spatial distribution of activities at a given point in time affect the location conditions for new activities.

The appearance of centripetal forces can be described by means of agglomeration effects. Weber (1909) understands an agglomerative factor as the advantage that results out of the existence of a certain density of a certain land use at a location. Agglomeration effects influence individual economic profits (profit/benefits), but are controlled by other economic entities. They can be divided into localisation effects (Localisation Economies) and urbanisation effects (Urbanisation Economies), although in the following we cover just the urbanisation effects.

Urbanisation effects arise between different actors and between different activities and represent those benefits arising from the entire scope of activities of a region, such as a

widely differentiated employment market, a broad spectrum of public and private services, high-quality communications infrastructure, innovation climate and good connections with and accessibility to other cities (cp. Maier & Tödtling, 2002, p. 101). One of the major urbanisation advantages is the availability of technical infrastructure (Kramar, 2005).

However, in the selection of a location, actors are influenced not only by agglomeration effects but also by opposing deglomeration effects. They describe the negative effects accompanied with overly high density agglomerated areas. As the density increases so too does the cost-benefit ratio. The costs include for example an overloaded city centre with corresponding traffic problems such as traffic jams, noise pollution or high land prices. Agglomeration disadvantages set in from the point at which the costs begin to outweigh the benefits (see Fig. 2).

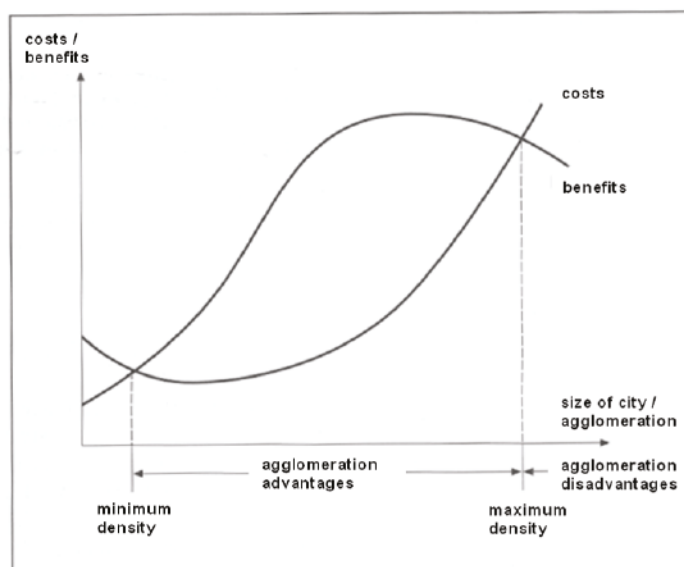


Fig. 2. Agglomeration advantages and agglomeration disadvantages as a function of costs/benefits. Figure from Kulke (2006, p. 242), source: Richardson (1976).

Based on the theoretical concept of the simulation model presented below, the costs curve corresponds to the agglomeration disadvantages. With increasing density the costs increase exponentially. The benefits curve represents the agglomeration advantages. The benefits continue to rise during ongoing aggregation until a saturation point is reached. As agglomeration continues, the usability of a location worsens, relative to its maximum advantageousness (see Fig. 2).

The difference between the benefit curve and the cost curve describes the probability of moving into or away from a location in the simulation model, as illustrated in Fig. 3. If the benefits outweigh the costs, the population density at the location will increase. On the other hand, if the costs outweigh the benefits, the population density at the location will be likely to decrease. During the development of a city (see chapter 2) the apex of the curve (the difference of costs and benefits) moves horizontally (see Fig. 3).

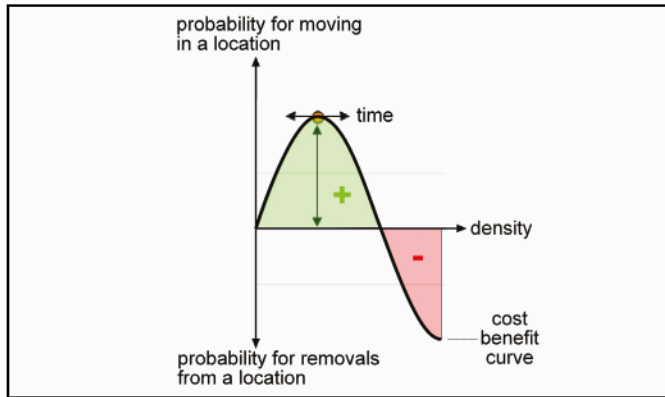


Fig. 3. Probability curve of moving into a location (green) or moving away from a location (red) as a function of density.

4. Modeling and simulation

4.1 Simulation concept

In the design of the simulation model, a key aim was to represent a complex dynamic system through a model that is as simple as possible. The model is intended to be used to explain the emergence of new qualities, to research the complex behaviour of an urban system and to predict future development trends.

Most of the common simulation models describe the growth of urban clusters as aggregation processes by means of DLA (diffusion limited aggregation) and DBM (dielectrical breakdown model) (Batty, 1991), which are both based on the principles of Cellular Automata (CA). In these models, parts (usually residential areas) are successively added to an existing structure by using a probability function. With these simple models, however, it is not possible to consider the change of the compactness of urban clusters over the course of time. For example, urban clusters exhibit rather compact structures during the cyclic phase of urbanism, that shift to become a more ramified structure in the transition from the cyclic phase of urbanism to the cyclic phase of suburbanisation. Furthermore, the specific phenomena for urban agglomerations, such as the restructuring of urban growth, the emergence of new growth zones or the coexistence of urban clusters can only be described by special model extensions (Schweitzer & Schimansky-Geier, 1994). A potential model expansion of the DBM principle exists for example in the formulation of several cell states and the corresponding transformation rules for the simulation of land-use patterns (White & Engelen, 1993). Schweitzer and Schimansky-Geier (1994) have demonstrated with their analysis of the maximum distance between a randomly selected point within the cluster and the next or the most distanced free space, that the shape of the urban residential area is not only controlled by the minimum distance to the city centre but by the minimum distance to the residential border as well. Simple DLA or DBM can be supplemented with such development rules (Schweitzer & Steinbrink, 2002).

Cities are complex systems and follow the principles of self-organization in their development. This hypothesis is a condition for the computer-based simulation model presented here. With the help of this model the occurrence of different phenomena in the

development of residential areas on an urban scale will be reviewed. The underlying analysis of the historical development of the city of Vienna from 1888 to 2001 (Müller, 2005) includes the general hypothesis that improving the accessibility of public transport systems and the supply through technical infrastructure systems leads to an increase in population density. This hypothesis will be verified using the model presented here. Within the model, the population density depends also on the individual weighting of a residential location in relation to the existing population density. This weighting will be based on the probability curve as shown in Fig. 3.

Fig. 4 provides an overview of how the presented hypotheses can be derived from the theoretical basis (see chapter 2 and 3) and how they can be verified using the simulation model described here.

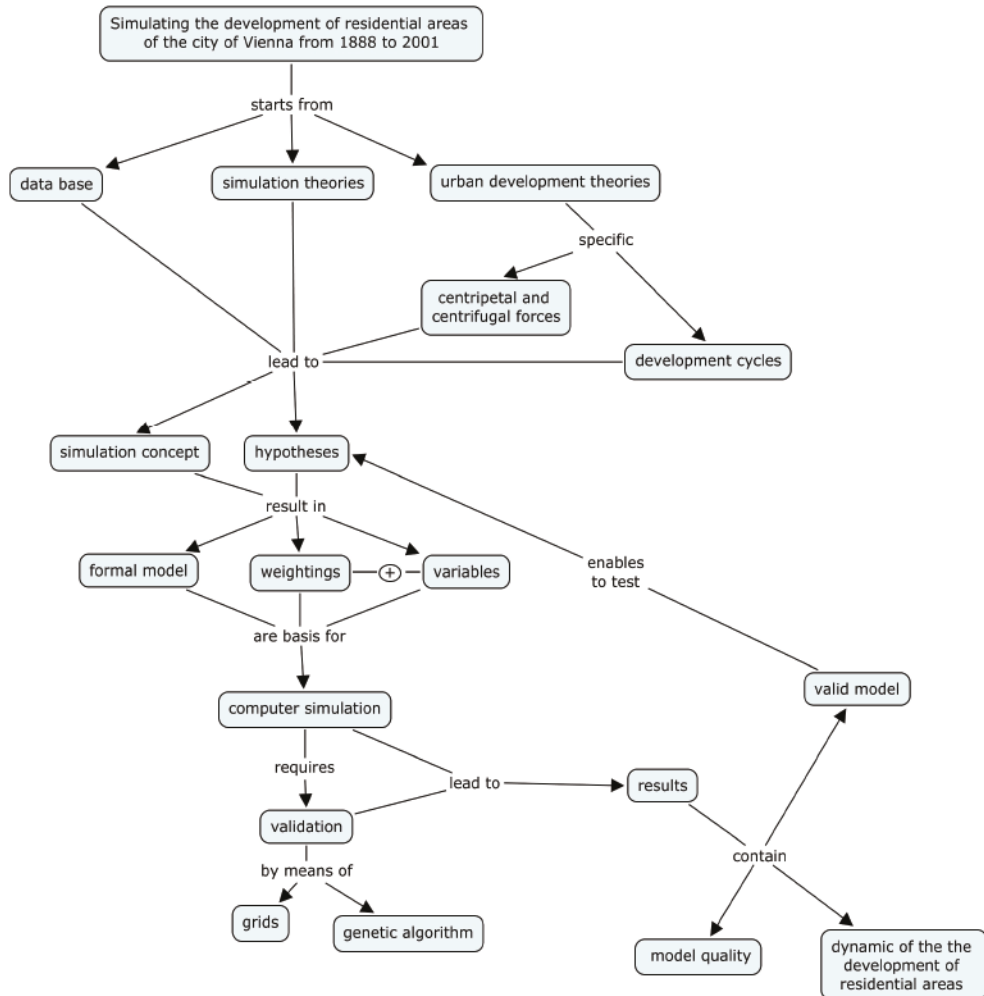


Fig. 4. Overview of simulation concept.

4.2 Data basis and variables

The data basis for the modelling is the analysis of the historical development of the city of Vienna from 1888 to 2001 based on information available from five points in time (1888, 1918, 1945, 1971, 2001). The indicators population density, accessibility of public transport systems and distance to technical infrastructure systems have been reviewed and evaluated.

- *Population density as a measure of the intensity of residential use:* Böventer (1979, p. 17) considers private households as seekers of residences on the one hand and as users of facilities, technical infrastructure and public transport systems on the other. The population density is defined as the sum of inhabitants per quarter hectare.
- *Accessibility of public transport systems as a parameter for exchange and interaction opportunities:* the accessibility of public transport systems is measured as the number of public transport stations within a maximum distance of 750 metres per grid-cell. The location of stations has been calculated using an approximation method, whereby the accessibility of public transport systems has not been determined as a distance dependent function but rather according to distance limits. Consequently the public transport stations are summarised as the public transport opportunities within a certain distance range or zone (Meise & Volwahren, 1980, p. 129).
- *Supply with technical infrastructure as a measure of local facilities of residential areas:* Technical infrastructure systems include among other things water, sewage, gas, electricity and district heating. The supply with technical infrastructure systems is defined as the density of infrastructural facilities within a radius of max. 1000 metres per grid. Each part of technical infrastructure route that intersects an urban grid cell is a potential infrastructural opportunity for the grid cell under consideration and the surrounding grid cells within a distance of 1000 metres. Thus, for each grid cell of the city of Vienna one can say: The more parts of technical infrastructure routes available in a radius of 1000 metre, the sooner the grid cell is supplied by the appropriate technical infrastructure system. By reaching the maximum value the grid cell under consideration is regarded as completely supplied. The lower the supply density with infrastructural facilities, the greater the potential for future expansion of the technical infrastructure system to achieve full supply.

The data of the three indicators were collected by dividing the city of Vienna into an area of 649×519 grid cells of 50 metre edge lengths.

4.3 Simulation

In the simulation model the cell grid of the CA represents the spatial structure of the city. The resolution of the basic grid corresponds to the configuration that was used when ascertaining the data (see section 4.2). Fig. 5 shows the principle for the simulation of the settlement spreading, which is based on the interaction of a so-called potential field and the development (settlement) of an individual area (Batty, 2005, pp. 105-150). In the following, a potential field is used for the calculation of the population density at a location (a grid cell) taking into consideration the population density in the neighbouring locations (neighbouring cells). Which areas of the potential gradient will be settled with a certain probability in the next time step, can be defined using the potential field principle.

For our investigation two potential fields are initiated. The first one represents the population density and the second one is a combination of the supply with infrastructural equipment and public transport accessibility. At each of the five points in time (1888, 1918,

1945, 1971, 2001) the potential of a cell is derived from the processed data on population density, technical infrastructure and public transport accessibility. By linear interpolation between these points in time, the growth rates of the population change as well as the extension of the infrastructural facilities per time step are specified. The location choices of settlement-agents take place endogenously and are represented in the following by a probability function.

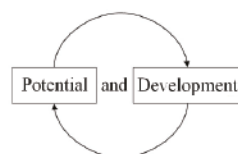


Fig. 5. Interaction of potential and development.

The cells of the model landscape can either be empty or populated with a certain density. With each time step – one month in the present simulation – a certain number of settlement-agents, depending on the growth rate, decide based on their current location preference, which locations they will settle or leave considering the potential values. Afterwards, taking into account the last settlement-activity, the potential values of all cells for the next time step are recalculated.

4.4 Weightings

For the simulation model we use various weightings for several indicators. Based on the statistical analysis of Müller (2005), the weightings for accessibility of public transport systems or for the supply with technical infrastructure are gathered from the results of individual regression analyses. These regression analyses explain at an examined point in time the population density at a specific rate through the accessibility of public transportation and availability of technical infrastructure. The standardised regression coefficients are used as weightings of the particular points in time in the simulation (Table 1 and Table 2). These weightings represent the percentages used to derive the supply potential field from the accessibility of public transportation and technical infrastructure (see equation 1).

	accessibility of public transport facilities	technical infrastructure
1888	50%	50%
1918	51%	49%
1945	26%	74%
1971	48%	52%
2001	58%	42%

Table 1. Weightings of the accessibility of public transportation and technical infrastructure.

The factor technical infrastructure is made up from the variables water, sewage, gas, and electricity. In a factor analysis it can be proven that these variables can be explained by the factor technical infrastructure (Müller 2005). The specific indicators of technical infrastructure (water, sewage, gas, and electricity) from the factor loadings out of the factor

analysis are available for the weightings at the current point in time and can be used in the simulation model (see equation (1)).

	gas	water	canal	electricity	district heating
1888	24%	37%	39%	0%	-
1918	25%	24%	27%	24%	-
1945	26%	23%	26%	24%	-
1971	25%	25%	25%	25%	18%
2001	25%	25%	25%	25%	15%

Table 2. Weightings of the indicators for technical infrastructure.

4.5 Formal model

For the formal representation of the model we agree on the following conventions. A cell of the CA is indicated with $H = \{1, 2, \dots, H_N\}$. The domain of the variable population density per cell D^H is normalised, i.e. scaled on the interval between 0 and 1 ($D^H = [0, 1]$). In the same way the domains of the supply variables per cell are normalised and indicated as follows: District heating $F^H = [0, 1]$, gas $G^H = [0, 1]$, sewage $K^H = [0, 1]$, water $W^H = [0, 1]$, electricity $S^H = [0, 1]$, public transportation system $O^H = [0, 1]$. From a summary of the supply variables the values of the supply potential field ($V^H = [0, 1]$) result:

$$V^H(t+1) = \omega_F \cdot F^H(t) + \omega_G \cdot G^H(t) + \omega_K \cdot K^H(t) + \omega_W \cdot W^H(t) + \omega_S \cdot S^H(t) + \omega_O \cdot O^H(t). \quad (1)$$

The factor ω indicates the fraction of a supply variable on the supply potential field V^H . The values of a supply variable for a cell H at time step t are calculated by linear interpolation of the collected data at the specified points in time and are taken as exogenous influence variables for the model. The variable for the population potential field is indicated with $P^H = [0, 1]$. The potential values are calculated with the following equation:

$$P^H(t+1) = \frac{1}{5} \cdot \left(\sum_{B \in U(H)} P^B(t) + D^H(t) \right). \quad (2)$$

The index B indicates a cell of the subset $U(H)$, which consists of the four directly neighbouring cells of a considered cell H without the considered cell H itself. The potential field is recalculated after each time step. The rates for positive R_p and negative R_n population growth results from the sums of the differences of the population densities of the individual cells between two points of time:

$$r^H = M^H(T+1) - M^H(T). \quad (3)$$

The points of time of the data assessment are declared with $T = \{1888, 1918, 1945, 1971, 2001\}$. The time steps between these points in time are indicated with t . In our case a time step covers one month. In contrast to the normalised population density D , the absolute population of a cell is indicated with M^H . The absolute change of a cell's population density between the points in time of data collection is declared with r^H . The growth rates R at a time step t depend on the scaling of model time, which is defined by the parameter c :

$$R_n(t) = \left(\sum_H r^H \mid r^H, r^H < 0 \right) / c \cdot ((T+1) - T). \quad (4)$$

$$R_p(t) = \left(\sum_H r^H \mid r^H, r^H > 0 \right) / c \cdot ((T+1) - T). \quad (5)$$

The differences of the population densities r^H are added to the negative growth rates (R_n), if $r^H < 0$, and otherwise, if $r^H > 0$, r^H is added to the positive growth rates (R_p). If the model time is scaled to month, for the denominator in equation 4 and 5 the value $12 \cdot (1918 - 1888) = 360$ results for the first simulation period.

At which cells the population density rises or falls is defined by assessment curves for moving away from or moving into individual inhabitants. These curves indicate which areas are evaluated as the least attractive ones by the current inhabitants due to the density and supply values of these locations and therefore that a decline in population is probable, or which areas are evaluated as the most attractive ones by the current inhabitants for the same reasons and therefore that an increase of population is probable. The probability ρ for moving away from a cell or moving to a cell of an inhabitant results from the assessment functions ρ_D for the population density and ρ_V for the supply potential of a cell:

$$\rho^H(t) = (\rho_D^H(t) + \rho_V^H(t)) / 2. \quad (8)$$

Following Krugmann (1996), the assessment function for the population density (ρ_D) results from a trade-off between two exponential functions for calculation of a centripetal F_{petal} and a centrifugal F_{fugal} force:

$$\rho_D^H(t) = F_{petal}^H(t) - F_{fugal}^H(t). \quad (7)$$

To simplify equation 7 we use the beta distribution¹ for the calculation of ρ_D which allows one to model an assessment curve with 2 instead of 4 control parameters. This will be relevant in chapter 6 where the search for optimal parameter settings is described. Since we assume different assessment curves for moving away from a cell or moving to a cell, equation 6 is to be distinguished in two cases, which in the following are indicated with ρ_{D_out} for the moving-away-from-curve and with ρ_{D_in} for the moving-to-curve:

$$\left. \begin{aligned} \rho_{D_out}^H(t) &= \frac{1}{B(p_{out}, q_{out})} \cdot x^{p_{out}-1} (1-x)^{q_{out}-1} \\ \rho_{D_in}^H(t) &= \frac{1}{B(p_{in}, q_{in})} \cdot x^{p_{in}-1} (1-x)^{q_{in}-1} \end{aligned} \right\}. \quad (8)$$

¹ cf. Wikipedia: http://en.wikipedia.org/wiki/Beta_distribution (last visited at 09.07.2008)

The course of the assessment curve² is defined by means of the parameter p_{out} and q_{out} as well as p_{in} and q_{in} . The term $B(p, b)$ indicates the beta function respectively for moving away from a cell and for moving to a cell:

$$B(p, q) = \int_0^1 x^{p-1} (1-x)^{q-1} dx. \quad (9)$$

We will deal with the concrete significance of the assessment curves in the following chapter 5.3. The assessment function for the supply-potential is a positive linear function to the supply-potential of a cell:

$$\rho_V^H(t) = V^H(t). \quad (10)$$

A location is regarded as being more attractive, the better its supply is. Based on the probability ρ , which is calculated for each cell, the decisions of the inhabitants where they move away from and where they move to can be made with the help of the roulette wheel method (Goldberg, 1989, S. 231). To choose a value ρ from the H_N values $\rho^1, \rho^2, \rho^3 \dots \rho^{H_N}$, the size of each probability value is specified by its weighting (size of the slot). At roulette wheel selection this weighting is indicated as selection probability w^H and can be calculated by dividing each value by the sum of all values:

$$w^H = \rho^H / (\rho^1 + \rho^2 + \rho^3 + \dots + \rho^{H_N}) \quad \text{oder} \quad w^H = \rho^H / \sum_H \rho^H. \quad (11)$$

To choose a value H the following algorithm is executed:

- a. generate a random value z between 0 and 1
- b. set $sum = 0$
- c. for $H = 1$ to H_N do
 - begin
 - $sum = sum + w^H$
 - if ($sum \geq z$) Then return H
 - end.

The randomly chosen value z corresponds to the position of the roulette ball and step c) tests in which slot H it ends up.

4.6 Computer program

The formal model presented in the previous chapter 5.2 was implemented in the programming language Delphi as a stand-alone simulation program for windows. Its basic functions are shown in Fig. 6.

The main area of the program window (Fig. 6, A) contains the graphic output, the visual representation of the various values of the separate variables which are saved per cell such as, for example, the population and supply density as well as the rates of moving away from or moving to. The darker the cells are (Fig. 6, A), the higher the density values. The green and red marked cells represent locations that inhabitants move to or move away from. The menu bar on the right-hand side of the program window is arranged in three register cards.

² An interactive animation of the beta distribution is available at:
<http://www.uni-konstanz.de/FuF/wiwi/heiler/os/vt-beta.html> (last visited at 09.07.2008)

Fig. 6, B shows the weightings for equation 1 and Fig. 6, C the parameters for the beta distribution of equation 8 for the definition of the course of the assessment curve for moving away from and moving to. The diagram shows the corresponding assessment curves for moving away from (red curve) and moving to (green curve).

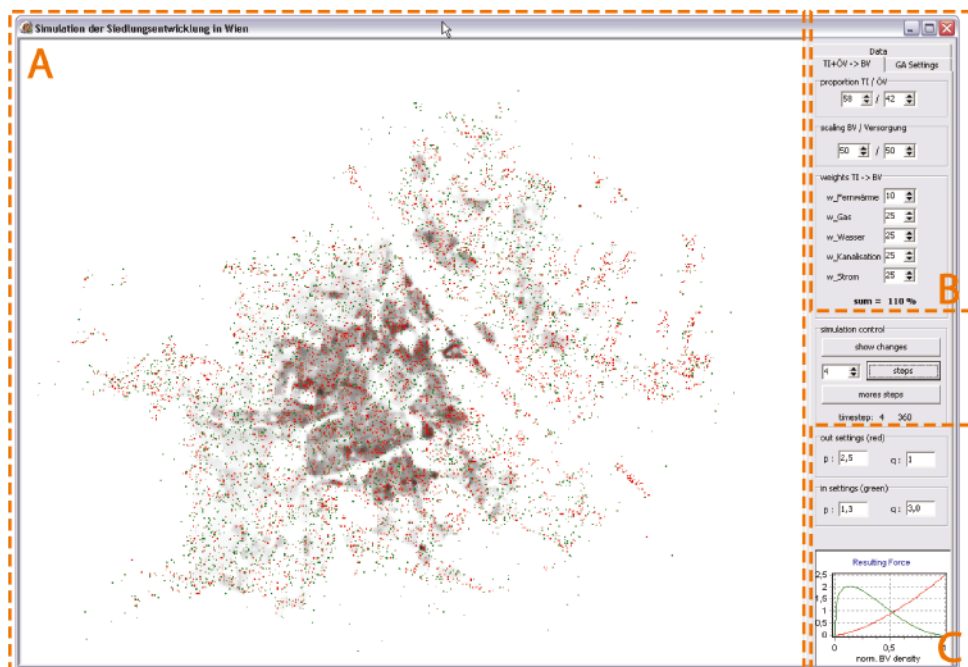


Fig. 6. Screenshot of the simulation program. Area A shows the population density in the year 2001. The considered period for moving away from or moving to in 5 time steps.

The calculation of the rates R_n and R_p for the population growth (see equations 4 and 5) are shown in Fig. 7. The first row (delta BV) declares the absolute population growth in the

Changes				
	1888-1918	1918-1945	1945-1971	1971-2001
delta BV	648957	-288376	-74636	-69578
total rate/t	1803	-890	-239	-193
+ BV	1095641	725661	584335	322985
+ rate/t	3043	2240	1873	897
- BV	-446571	-1014279	-658781	-392638
- rate/t	-1240	-3131	-2111	-1091
steps	360	324	312	360
+ big_BV	1095641	725661	584335	322985
+ big_rate/t	3043	2240	1873	897
- big_BV	-446571	-1014279	-658781	-392638
- big_rate/t	-1240	-3131	-2111	-1091

Fig. 7. Screenshot of the rates of change.

periods of the corresponding columns. The second row (total rate/ t) comprises the interpolated absolute monthly growth rate per time step t . The third row (+BV) shows the absolute number of influxes in a period. In the fourth row (+rate/ t) the influxes per time step $R_p(t)$ are declared. The fifth row (-BV) shows the absolute number of movements away from locations in a period. In the sixth row (-rate/ t) the movements away from locations per time step $R_n(t)$ are declared. In the seventh row (steps) the number of time steps t (in months) of the respective period are declared.

5. Validation

The validity of a simulation run is judged at the end of a period by means of the population distribution. One cannot expect the simulated data and surveyed data of population density to correspond at a detailed level (e.g. per cell) because of the abstraction for example of topographic, social and economic conditions. Consequently, the correspondence between the population densities is measured in radial areas, which are defined by the distance to the city centre (centre of mass). The position of a cell H in a coordinate system can be defined as vector v^H . The centre of mass Z can be calculated by forming the sum of the weighted vectors and dividing it by the total number of settled cells (Schweitzer & Steinbrink, 2002). The weight of a vector corresponds to the population density of a considered cell:

$$\rho_V^H(t) = V^H(t). \quad (12)$$

The population density is measured in concentric circles around the centre of mass Z (Fig. 8, left). Now the population differences between real data and the results of the simulations inside the measure rings are calculated. The population difference summed up for all rings represents the quality of the model. The model quality is best if the absolute sum of the population difference is as low as possible. Based on the real data a simulation run starts at $T-1$ and ends at T . Therefore the model quality refers always to the period under consideration.

Fig. 8 (right-hand side) shows the various measure charts. Both charts above show the measurements of fractal dimensions. Diagram A represents the differences of separate measure cells. Diagram B shows the measure rings, which are used for the validation of the simulation. The blue graph represents the differences per ring and the red graph shows the sum of the differences.

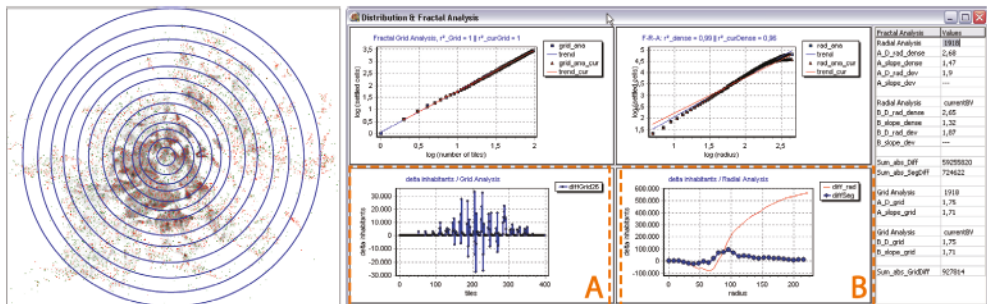


Fig. 8. Validation method.

The result of the simulation respectively the quality of the model that results after a simulation run from $T-1$ to T depends mainly on the settings of the control parameters p and q for the definition of the assessment curve by means of equation 8. The values for the control parameters are kept constant during a simulation period (from $T-1$ to T). The challenge for the validation of the present model was to search for the optimal settings for these control parameters for each considered period. Our aim was to achieve the best possible model quality and therefore validity for each period.

5.1 Genetic algorithm

To find the optimal settings for the control parameters a genetic algorithm was used. The base forms the structure of the so-called simple genetic algorithm (SGA) according to Goldberg (1989, p. 69), which was adapted for the present challenges. The initial population for the SGA are 32 individuals that consist of a randomly generated chromosome of 28 binary numbers. In these chromosomes the four control parameters for the assessment curves of equation 8 are coded (Fig. 9). For each individual, a simulation period is run with the corresponding control parameters and the corresponding model quality is calculated. This quality represents the evaluation criteria for the suitability of a control parameter setting and can therefore be used for the fitness definition of an individual.

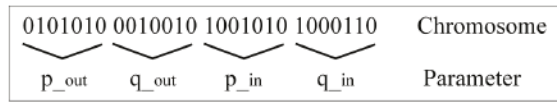


Fig. 9. Coding of the parameters in a chromosome.

After the simulation period has been run through for each of the 32 individuals and a fitness value has been assigned to each of them, a new generation of individuals is generated. The probability for the reproduction of an individual is defined by its fitness value. The fitness values are scaled in such a way that the individual with the highest value is selected with twice the probability as an individual with an average fitness. The individual with the lowest fitness has no chance of reproduction.

The crossing rate indicates the probability that the selected individuals are crossed or copied without a change into the new generation. Crossing means the mixture or recombination of two (parent) chromosomes (Fig. 10). At the present SGA we use a crossing rate of 80%.

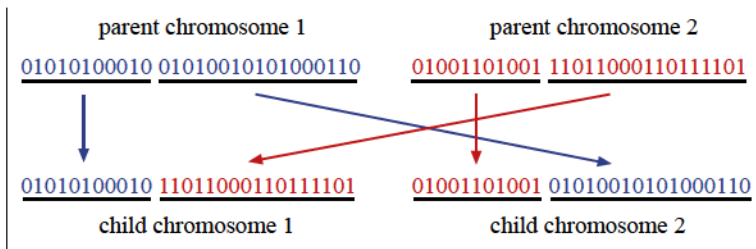


Fig. 10. Generation of child chromosomes by crossover of the parent chromosomes.

The mutation rate indicates the probability for the inversion of a randomly selected binary number of a chromosome, e.g. from 0 to 1. At the present model we have set the mutation rate to 1%. Therefore, on average, 1 out of every 100 binary numbers mutates.

5.2 Grids

With the available data we have to consider all possible locations (cells) of the cell landscape for moving to a cell or moving away from a cell of inhabitants per time step t . The cell landscape includes $649 \times 519 = 336,831$ cells. For the first period from 1888-1918 per time step t , 3,043 influxes and 1,240 effluxes have to be calculated. This results in 4,283 calculations per t , which in turn means 1,542,212 calculations for the whole period. With current computers, the simulation of this period lasts approximately 4.5 hours. For one simulation per period this would be an acceptable time. It becomes problematic, however, when in order to search for the optimal control parameter settings many hundreds of periods have to be calculated (see section 6.1), since the overall computing time can quickly add up to several weeks. For this reason we vary the resolution of the cell landscape denoted by the grid in the following. To change the resolution, the existing population data of several 50×50 metre cells are merged in a larger cell. The original cell raster with an edge length of 50 m per cell is denoted as grid 1. At grid 5, 5×5 original cells are merged to a cell with a side length of 250 m and at grid 10, 10×10 original cells are merged to a cell with a side length of 500 m and so on (Fig. 11).

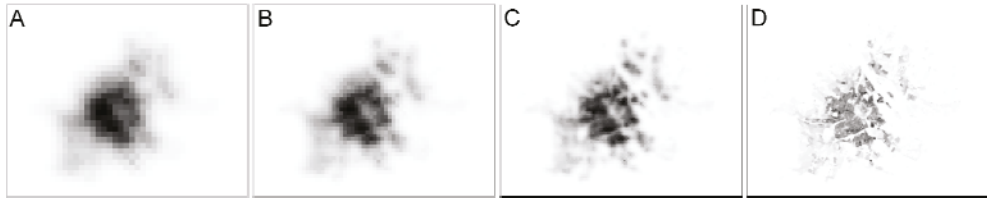


Fig. 11. Representation of the population density with various grid resolutions. A) grid 15, B) grid 10, C) grid 5, D) grid 1.

The use of various grids allows a comparison of simulation results at different degrees of data precision. Does it make any difference whether we have population density data for grid 20 or grid 1 and do the resulting optimal control parameters differ only slightly? This question is answered in the description of the results in the next chapter.

In the search for optimal control parameters, as a first step, the optimal values for the control parameters for the coarsest grid (grid 20) are calculated. In a second step the results of these calculations are integrated in the initial population for the next finer grid (grid 15) to calculate the optimal control parameters for this grid. These steps are repeated for grid 10, grid 5, and grid 1. We act on the hypothesis that the transfer of the optimal values from the coarser to the finer grids decrease the necessary number of generations and with this the overall computing time necessary for the search by means of a genetic algorithm. Whereas up to 100 generations were calculated at grid 20, the search for the optimal values was aborted after 5 generations per period at grid 1. Nevertheless at grid 1 a model quality between 75% and 90% could be achieved for all periods (Fig. 12).

6. Results and conclusions

6.1 Model quality

In chapter 6 the model quality was defined as totalised absolute difference between simulated and real population density over the individual measuring rings. The best model quality (100%) is achieved when simulated and real population densities match exactly. The model qualities achieved for the various periods with regard to the different grids are shown in Fig. 12.

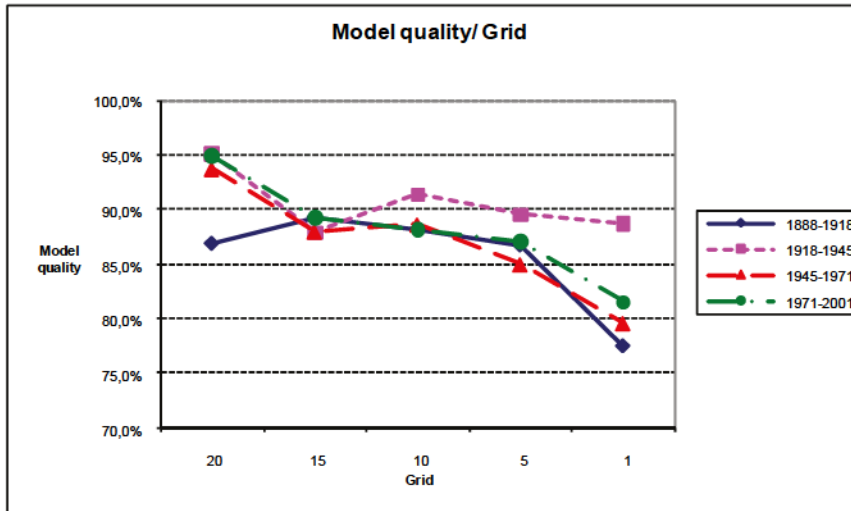


Fig. 12. Model quality (ordinate) at optimal values for the four control parameters for the various periods depending on different grids (abscissa).

In principle for all periods the achieved model quality is very good with values over 85% for grids greater than 1. There are several possible reasons for the quite obvious decrease in quality for all periods for grid 1: firstly, the measuring method used and its accuracy were adapted for each grid. Secondly, the small number of calculated generations for grid 1, since the optimal parameters do not agree with those of the coarser grids (see Fig. 13) and therefore could probably be optimised by calculating further generations. Though, this was not possible within this project for reasons of time. Thirdly, the limiting of the maximum value (upper bound) for the four control parameters at 12.7 has possibly led to some distortions, since the calculated optimal values for q_{in} at T_3 and T_4 have already reached this maximum (see Fig. 13).

6.2 The valid model

The most important results of the project are the optimal values for the four control parameters p_{out} and q_{out} as well as p_{in} and q_{in} that are found by means of the genetic algorithm. The various coloured bars in Fig. 13 show the values of the control parameters, which have led to the best results at the different grids (best model quality).

In Fig. 13 it is clearly visible that at each period some parameter values change only slightly while others change very strongly if we compare the different grids. For example the values for q_{out} and p_{in} at T_1 in Fig. 13, character A are relatively similar from grid 20 to grid 1, whereas in the same period the values for p_{out} vary quite significantly from grid 20 to grid 1. The variations of individual parameters lead to relevant differences in the course of the assessment curves for moving away from or moving to. These differences in the course of the assessment curves become apparent if we compare, for example, the mean values of the control parameters for the four periods with the optimal values for grid 1, as shown in Fig. 14 to Figure 16.

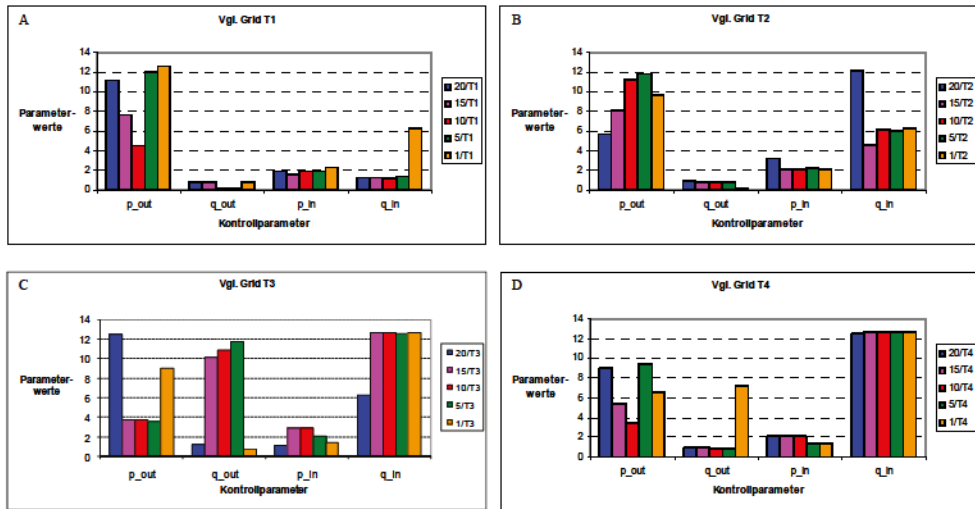


Fig. 13. Comparison of the optimal control parameters per time step T at different grids (grid 20 to grid 1 in the legend). A) T1=1888-1918, B) T2=1918-1945, C) T3=1945-1971, D) T4=1971-2001

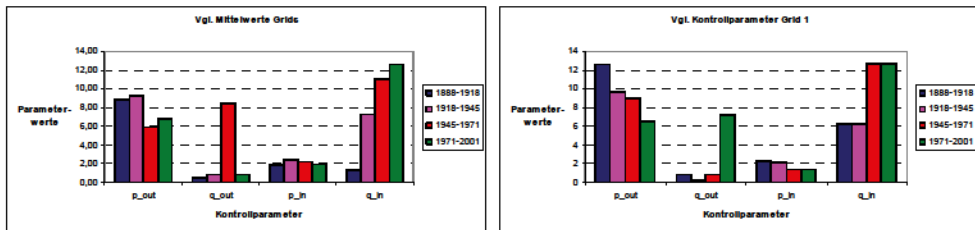


Fig. 14. Development of the control parameters. Left: Mean values of the optimal values of the four control parameters for different grids per period T (cf. Fig. 13). Right: Optimal control parameters for grid 1.

The diagram on the left in Fig. 14 shows the mean values of the optimal values for the four control parameters p_{out} , q_{out} , p_{in} and q_{in} for different grids per period T1 to T4 with strongly divergent values for q_{out} in the third period and for q_{in} in the course of all considered periods (T1 to T4). At the optimal parameter values for grid 1 in Fig. 14 right, the values for q_{in} vary very little from each other in both first and in both last periods. On the other hand, during the same period the values for p_{out} vary considerably from each other. For q_{out} the strongest variation is in the period from 1971 to 2001.

Using the mean values (Fig. 14, left), the assessment curves for moving to and moving away from are created, which are shown in Fig. 15. From an interpretation of these assessment curves we can conclude that, for example, at T_3 in Fig. 15, character C, the population has moved from the outer urban districts to rural areas, whereas at T_4 in Fig. 15, character D, the population redistribution was from the city centre to rural areas.

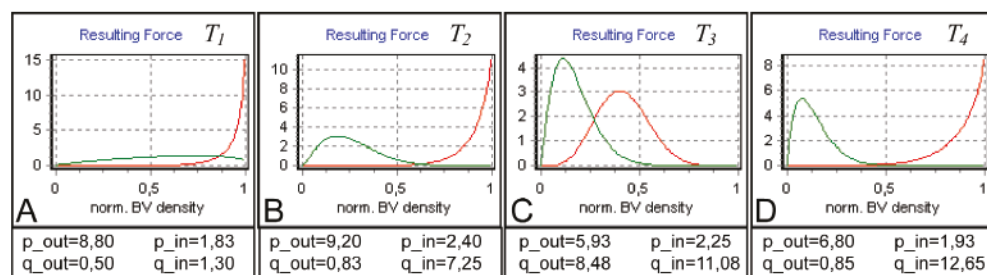


Fig. 15. Assessment curve for the mean values from Fig. 14, left. A) T1=1888-1918, B) T2=1918-1945, C) T3=1945-1971, D) T4=1971-2001

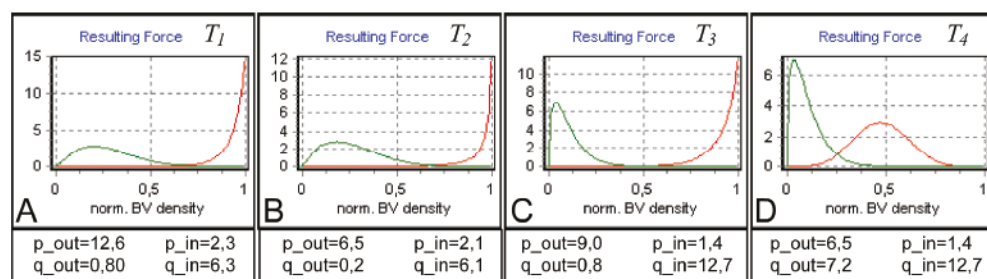


Fig. 16. Assessment curve for the best values for grid 1 from Figure 13. A) T1=1888-1918, B) T2=1918-1945, C) T3=1945-1971, D) T4=1971-2001

The assessment curves for grid 1 in Fig. 16, character C and character D, have to be interpreted just contrarily. The curves show that the population moves from outer urban areas to rural areas at T_4 . However, a comparison of the green moving to curves of the diagrams in Fig. 15 and Fig. 16 we can observe that they both show a similar course. The interpretation of the moving to curves is essentially the same in Fig. 15 and Fig. 16.

If we consider the differences described between the graphs in Fig. 15 and Fig. 16 we can conclude on the one hand that the exactness of the data (in the present application represented by the cell size) has a considerable effect on the results of the simulation. Consequently, the significance of the results of the present study are relative, since they are related to the population and supply data from a 50×50 m grid (see chapter 4.2)

However, if we look at the similarities between the assessment curves in Fig. 15 and Fig. 16 we can conclude on the other hand that even with a much coarser grid and correspondingly less accurate data, the values for the control parameters can be calculated with sufficient exactness and acceptable significance.

6.3 The dynamics of settlement development

Apart from the development of a valid model we could have shown the suitability of the four control parameters p_{out} and q_{out} as well as p_{in} and q_{in} (see equation 8) as formal representations of the individual assessment of a residential location concerning the existing population density. With the simulation, the hypotheses (see chapter 4.1) could be supported whereby an improvement in the accessibility of public transport facilities and the proximity to technical infrastructural facilities causes an increase in population density. The assessment curves resulting from the control parameters of grid 1 allow a visual interpretation of the calculated values for these parameters.

Considering the green curves in the diagrams in Fig. 16 it can be observed that the preferred locations for influxes have shifted more and more from the outskirts of the town (at T_1 and T_2) to the rural areas (at T_3 and T_4). The population settles increasingly in those areas with lower population density. As a result the suburbanisation process (see chapter 2.2) in Vienna from 1888 to 2001 can be identified in the shift of the green curves maximum in the diagrams in Fig. 16. Suburbanisation is most clearly visible between the Second World War and the last considered point in time. This process is also apparent in the consideration of the red curves in Fig. 16. There we see that the movements away concentrates first on the city centre (from T_1 to T_3), shifting only in the last period to the outer urban districts (at T_4). The overall illustration of moving to and moving away from (Fig. 17) over the course of time confirms our interpretation of the assessment curves.

At the first point in time T_1 the migration from the exurban fringe already becomes apparent. The population is moving to mainly sparsely settled areas – at T_1 primarily beyond the former town's fortification. Over the course of time – until the last point in time T_4 – the population moves increasingly away from the centrally-located densely-populated urban areas. The progressive suburbanisation process becomes more and more obvious since the population increasingly prefers sparsely populated or empty areas as residential areas. We can observe an overall spread of the settlement towards as well as beyond the city limits.

In chapter 3 we have already indicated that the position of the maximum and the form of an assessment curve for movements to and away depends on various economic, social, cultural and ecological general conditions. For the present simulation model the form of the curves was determined only for four periods (Fig. 16). The assumption that the assessment curves dynamically change over the course of time is more realistic. According to this, starting at T_1 , a curve of a diagram in Fig. 16 would gradually turn into the same coloured curve at

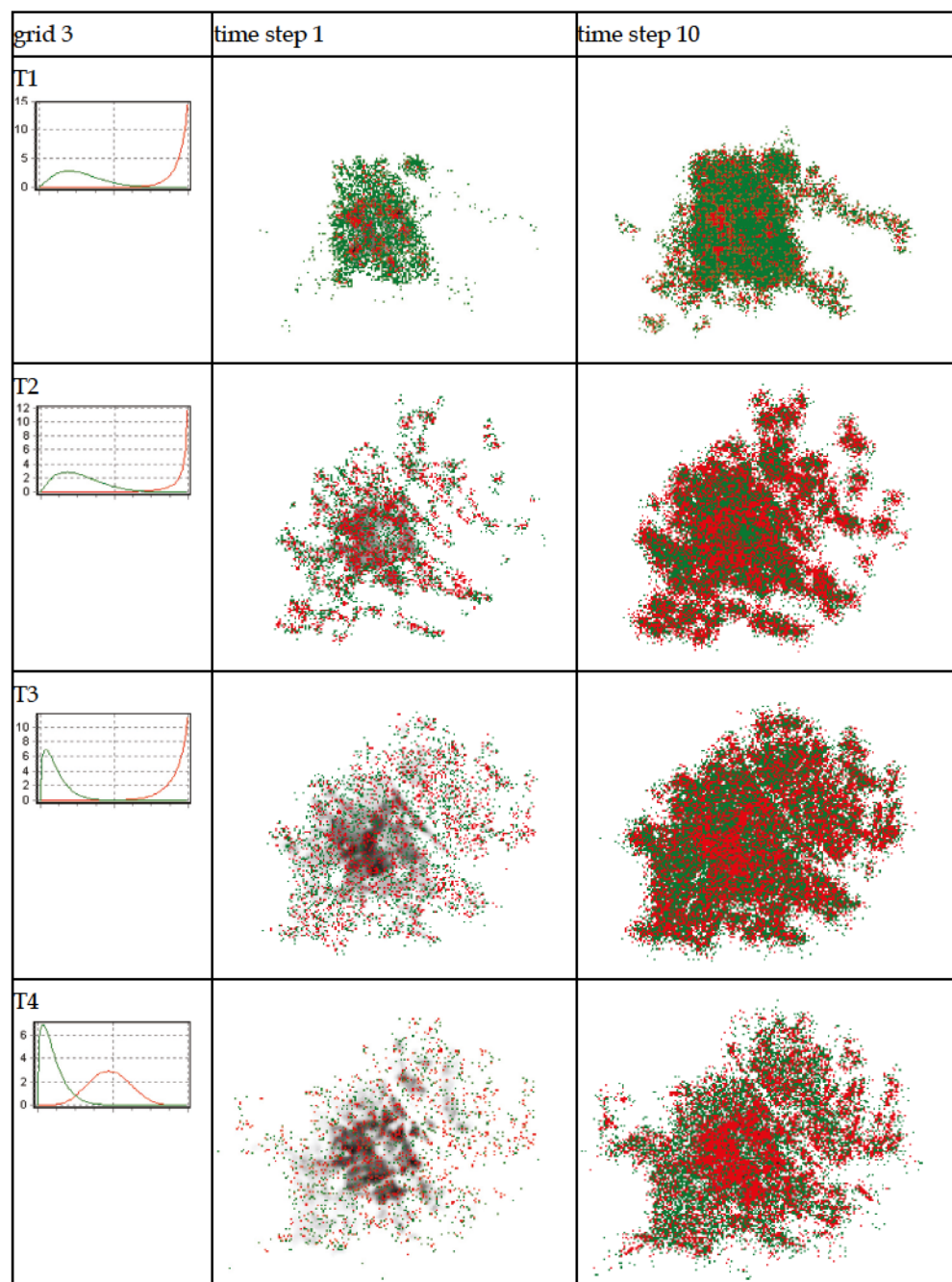


Fig. 17. Overall illustration of the movements to and movements away after one and ten time steps (second and third column) for different periods (rows T1 to T4)

point T in time. The concept for the change of the assessment curves over the course of time provides a robust explanatory approach for describing the basic processes which are responsible for the change of settlement structures.

7. Summary and outlook

The relatively high model quality indicates that we have achieved a high validity with the model shown here for simulating the population distribution of the city of Vienna from 1888 to 2001. With the help of four control parameters for the definition of the assessment curves we have developed a sensible and robust concept for the explanation of the driving forces of urban development processes. Through the exogenous definition of the infrastructural supply of a location we could show that the development of the population density can be essentially regulated by infrastructure investments. Furthermore the suburbanisation processes, which have taken place over the course of the considered period in the city of Vienna, can be explained by our model.

The simulation model was designed to be as simple and transparent as possible to enable a sensitivity analysis of the control parameters and through this to ensure the comprehensibility of the results. Only through the construction of comprehensible models will it be possible to improve the acceptability of simulations for political decision makers or urban and regional planners. Similarly, only models whose operating principles are described transparently can be improved further and further in the course of its usage. The next steps for the development of the model are to first explain the extension of the technical infrastructure and the spread of the public transport facilities on the basis of the development of the population density, and secondly to adapt the model in a way that the development of the population density can be circularly coupled with the availability of technical infrastructure and the spread of public transport facilities. By adding supplemental exogenous parameters such as, for example, topographical conditions, the significance of the model could be increased still further.

Furthermore, one could examine to what extent changes to the control parameter values in the present simulation can be connected with characteristic values of the economic and technological development of the city of Vienna during the investigated period.

The application perspective for the model at hand is to estimate the consequences of particular planning activities on the basis of restricted scenario models. Such a scenario could, for example, include planned infrastructure expansion or the effects of increasing transportation costs on the assessment curves and with this on the development of the settlement structure.

8. Acknowledgements

This research project was financed by the *Jubiläumsstiftung* of the city of Vienna.

9. References

- Batty, M. (1991). Generating urban forms from diffusive growth. *Environment and Planning A*, 23(4), 511-544.

- Batty, M. (2005). *Cities and Complexity: Understanding Cities with Cellular Automata, Agent-Based Models, and Fractals*. London: MIT Press.
- Behrens, K. C. (1971). *Allgemeine Standortbestimmungslehre*. Opladen: Westdeutscher Verlag GmbH.
- Böventer, E. (1979). *Standortentscheidung und Raumstruktur*. Hannover: Schroedel Verlag.
- Dangschat, J. (2007). Reurbanisierung – eine Renaissance der (Innen-)Städte? *Städtepolitik und Stadtentwicklung*, 3.
- Denkschrift über Grundsätze des Städtebaues. (1906). In R. Baumeister (Ed.), *Beiträge zum Städtebau: Association of German Architects and Engineers*.
- Fassmann, (2005). *Stadtgeographie I*. Braunschweig: Westermann
- Friedrichs, J. (1981). *Stadtanalyse: Soziale und räumliche Organisation der Gesellschaft*. Opladen: Westdeutscher Verlag.
- Goldberg, D. E. (1989). *Genetic Algorithms in Search, Optimization and Machine Learning* (1 ed.). Boston: Addison-Wesley.
- Kramar, H. (2005). Innovation durch Agglomeration: Zu den Standortfaktoren der Wissensproduktion. In D. Bökemann (Ed.), *Wiener Beiträge zur Regionalwissenschaft* (Vol. 20). Wien.
- Krugmann, P. (1996). *The Self-Organizing Economy*. Cambridge, Mass.: Blackwell.
- Kulke, E. (2006). *Wirtschaftsgeographie*. Paderborn: Schöningh Verlag
- Läpple, D. (2003). Thesen zu einer Renaissance der Stadt in der Wissensgesellschaft. In N. Gestring, H. Glasauer, C. Hannemann, W. Petrowsky & J. Pohlen (Eds.), *Jahrbuch Stadt/Region* (pp. 61–78). Opladen: Leske und Budrich.
- Lichtenberger, E. (1986). *Stadtgeographie: Begriffe, Konzepte, Modelle, Prozesse* (Vol. 1). Stuttgart: Teubner.
- Maier, G., & Tödtling, F. (2002). *Regional- und Stadtökonomie: Standorttheorie und Raumstruktur*. Wien, New York: Springer Verlag.
- Meise, J., & Volwahn, A. (1980). *Stadt- und Regionalplanung, Ein Methodenhandbuch*. Braunschweig: Friedr. Vieweg & Sohn Verlagsgesellschaft
- Müller, D. (2005). *Wien 1888 – 2001: Zusammenhänge der Entwicklung der technischen Infrastruktur- und ÖV-Systeme in den Siedlungsgebieten*. Wien: Peter Lang Verlag.
- Myrdal, G. (1957). *Economic Theory and Underdeveloped Regions*. London: Duckworth.
- Richardson, H. W. (1976). Growth Pole Spillovers: the dynamics of backwash and spread. *Regional Studies. The Journal of the Regional Studies Association*, 10(1), 1-9.
- Schweitzer, F., & Schimansky-Geier, L. (1994). Clustering of Active Walkers in a Two-Component System. *Physica A*, 206, 359-379.
- Schweitzer, F., & Steinbrink, J. (2002). Analysis and Computer Simulation of Urban Cluster Distribution. In K. Humpert, K. Brenner & S. Becker (Eds.), *Fundamental Principles of Urban Growth* (pp. 142-157). Wuppertal: Müller + Busmann.
- Van den Berg, L., Drewett, R., Klaassen, L. H., Rossi, A., & Vijverberg, C. H. T. (1982). *Urban Europe: A Study of Growth and Decline* (Vol. 1). Oxford: Pergamon Press.
- Weber, A. (1909). *Über den Standort der Industrie*. Tübingen: J.C.B. Mohr Verlag

- White, R., & Engelen, G. (1993). Cellular automata and fractal urban form: a cellular modelling approach to the evolution of urban land-use patterns. *Environment and Planning A*, 25(8), 1175-1199.